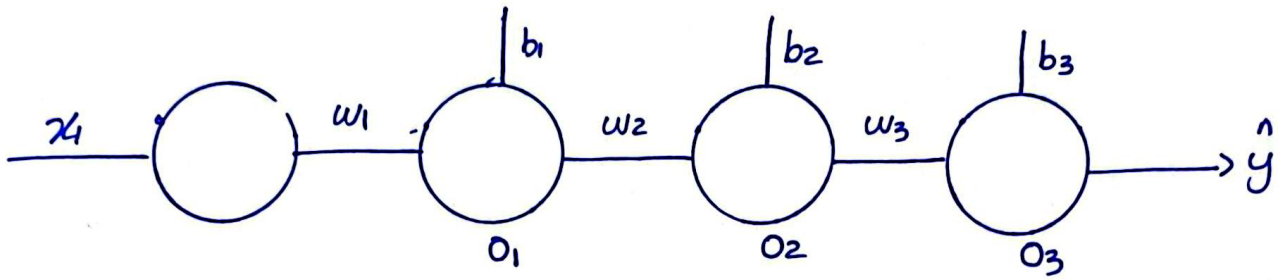


Exploding Gradient Problem



$$w_{\text{new}} = w_{\text{old}} - \alpha \frac{\partial L}{\partial w_{\text{old}}}$$

$$\frac{\partial L}{\partial w_{\text{old}}} = \frac{\partial L}{\partial O_3} \times \frac{\partial O_3}{\partial O_2} \times \frac{\partial O_2}{\partial O_1} \times \frac{\partial O_1}{\partial w_{\text{old}}}$$

$$\frac{\partial O_3}{\partial O_2} = \frac{\partial \sigma(z)}{\partial z} \times \frac{\partial z}{\partial O_2} \quad z = O_2 w_3 + b_3$$

$$= [0 - 0.25] \frac{\partial (O_2 w_3 + b_3)}{\partial O_2}$$

$$= [0 - 0.25] w_3 \quad \therefore w_3 \Rightarrow \text{high value}$$

Weight Initialization

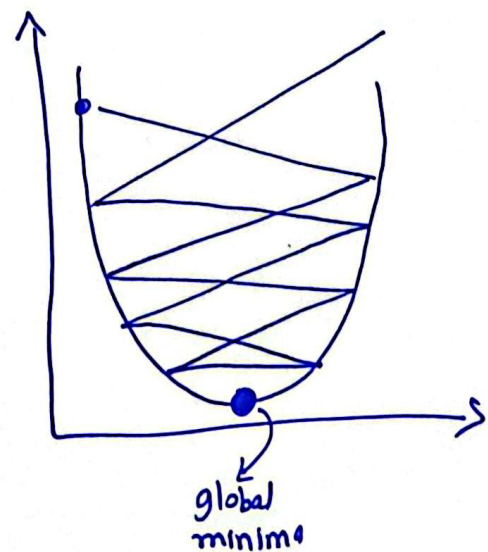


very high value



$$w_{\text{new}} \gg w_{\text{old}}$$

$$w_{\text{new}} \ll w_{\text{old}}$$



This issue can be fixed

with **Weight Initialization Techniques**